点推定における漸近有効性とその問題点(統計的決定方式の最適性の研究)

高木, 祥司

数理解析研究所講究録
第 723 巻 1990 年 15-23

点推定における漸近有効性とその問題点

大阪大学基礎工学部　$M2$　高木祥司

平成元年 11 月 16 日

1　序文

以前より，推定量の良さの測度として有効性の概念が考えられてきた．そして，この概念において，最尤推定量 (MLE) は最良な推定量であるか，という理論が様々な方法で研究されてきた．そこで，2 節でいくつかのよく知られた概念を述べ、3 節で、Bahadur の意味での漸近有効性の概念を述べ，その特長や問題点を例を用いて考えてみることにする．

2 記号と一般的な概念

以下で用いる記号を次のように定義する．
確率変数を　$X_1, X_2, \cdots$, 母数空間 Θ は開区間 ,
密度を $f(x \mid \theta)$, 尤度を $f(s \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$
対数密度を $l(x \mid \theta) = \log f(x \mid \theta)$, 導関数を $l'(x \mid \theta) = \frac{\partial}{\partial \theta} l(x \mid \theta)$
対数尤度を $l(s \mid \theta) = \log f(s \mid \theta)$, Fisher 情報量を $I(\theta)$, 統計的曲率を $\gamma(\theta)$ とする．ここで、

$$\mu_{ijk} = E(\frac{f'}{f})^i (\frac{f''}{f})^j (\frac{f'''}{f})^k$$

とすると、$I(\theta) = \mu_{200}$,

$$\gamma(\theta)^2 = \frac{1}{I^2}(\mu_{020} - 2\mu_{210} + \mu_{400}) - 1 - \frac{1}{I^3}(\mu_{110} - \mu_{300})^2$$

となり、この統計的曲率は、分布がどれ程指数族に近いかを表していて、指数族の場合に 0 となり、一般には 0 より大きい．
また推定量 T_n の Fisher 情報量を $I_T(\theta)$ とする．

Fisher は，情報量の損失で推定量の有効性を測ることを提案した (Fisher[5],[6])．
つまり，

$$(2.1) \qquad nI(\theta) - I_T(\theta)$$

の値が小さいほど T_n は良い推定量であるとし，MLE の漸近有効性を予想した．
以後の Rao-Efron の研究により，MLE の最適性に関する次の定理が成立する．(Rao[13],Efron[4])

定理 2.1　ある正則条件の下で

$$(2.2) \qquad \lim_{n\to\infty}\{nI(\theta)-I_T(\theta)\} \geq I(\theta)\gamma^2(\theta),$$

$$(2.3) \qquad Var(T_n) \geq \frac{1}{nI}+\frac{1}{n^2I}(\gamma^2+4\frac{\Gamma^2}{I})+2\frac{\dot{b}}{n^2I}+o(n^{-2}).$$

ここで，$b(\theta)$ はバイアス，つまり，$\frac{b(\theta)}{n}=E(T_n-\theta)$ で，$\dot{b}=\frac{d}{d\theta}b(\theta)$，
Γ は，T_n に依存しない数である．
また，$T_n=\hat{\theta}_n$ (MLE) の時，いずれも等号が成立する．

また、Rao は、適当な a,b が存在して、

$$\mid \frac{1}{\sqrt{n}}\frac{d}{d\theta}l(s\mid\theta)-a-b\sqrt{n}(T_n-\theta)\mid \to_P 0$$

ならば、Rao の意味で一次漸近有効であると定義した．この時、Fisher 情報量と次の様な関係がなりたつ．

定理 2.2　(Rao[12])
Rao の意味で一次漸近有効ならば、$\frac{I_T}{n}\to I$

また $\frac{d}{d\theta}l(s\mid\theta)-a\sqrt{n}-bn(T_n-\theta)-\lambda n(T_n-\theta)^2$ の分散を最小にする λ を選び、その分散の最小値 E_2 の大きさを二次の漸近有効性の基準と考えた．そして、

$$\lim_{n\to\infty}(nI-I_T)=E_2$$

を予想したが、Efron[4] によって否定された．

3　Bahadur による漸近有効性

Bahadur は一致推定量 T_n の良さを θ の $\varepsilon-$ 近傍外での集中確率の大きさ，つまり，

$$\alpha(T_n,\theta,\varepsilon)=\Pr(\mid T_n-\theta\mid>\varepsilon)$$

で測ることを提案し，この値が小さいほどよい推定量であると，考えた．しかし，この値を計算するのは難しく，正確に取扱うことができない．ましてや，最良なものは，容易には見つけることができない．そこで，この値が，一

般に，$n \to \infty$ の時，指数的に 0 に収束することより，

$$\beta(T_n, \theta, \varepsilon) = \lim_{n\to\infty} -\frac{1}{n} \log \Pr(|\, T_n - \theta \,| > \varepsilon)$$

の大きさを，一つの測度とする (Bahadur[1]). ここで，この量は，exponential rate または，inaccuracy rate と呼ばれ，この値が大きいほど良い推定量であると考える．そこで，この量を $\varepsilon \to 0$ で評価することにより，推定量の有効性を定義する．一般に次の定理が，成立する．

定理 3.1　T_n を任意の一致推定量とする．

(1) (3.1) $\limsup_{n\to\infty} -\frac{1}{n} \log \Pr(|\, T_n - \theta \,| > \varepsilon) \leq \inf\{K(\theta', \theta) : |\, \theta' - \theta \,| > \varepsilon\}.$

ここで, $K(\theta', \theta)$ は，Kullback 情報量を表わす．また，右辺は，Bahadur bound と呼ばれ，$B(\theta, \varepsilon)$ で表わす．

(2) 次の正則条件 (RC) の下で，以下の 2 つの式が，成立する．(Fu[7])
(i) すべての $\theta \in \Theta$ に対して，$K(\theta', \theta)$ は，δ 近傍 $N_{\theta,\delta}$ 内の θ' において，局所 convex 関数である．
(ii) すべての　$\theta \in \Theta$ に対して，適当な $N_{\theta,\delta}$ と，関数 $A(x, \theta)$ が存在して，すべての $\theta', \theta'' \in N_{\theta,\delta}$ において，$E(A^2) < \infty$, $|\, l'(x \mid \theta'') - l'(x \mid \theta') \,| \leq A \,|\, \theta'' - \theta' \,|$
(iii) すべての $\theta \in \Theta$ に対して，2 つの定数 $u = u(\theta) > 0, w = w(\theta) > 0$ が存在して，$0 < \varepsilon < u$ において，$\Pr\{l'(x \mid \theta + \varepsilon) > 0\} > 0$ かつ，$\Pr\{l'(x \mid \theta - \varepsilon) < 0\} > 0$ であり，しかも，$\phi(t, \theta, \varepsilon) = E[exp\{tl'(x \mid \theta + \varepsilon)\}]$ が，すべての $(t, \varepsilon) \in (-u, u) \times (-w, w)$ に関して有限である．
(iv) 偏導関数 $(\frac{\partial}{\partial t})(\frac{\partial}{\partial \theta})\phi(t, \theta, \varepsilon)$ が存在し，$(t, \varepsilon) \in (-u, u) \times (-w, w)$ に対して，連続である．
(v) 各 n と s に対して，MLE $\hat{\theta}_n$ は，$l'_n(s \mid \theta) = 0$ の，唯一の解である．

$$\text{(3.2)} \quad \limsup_{\varepsilon\to 0} \limsup_{n\to\infty} -\frac{1}{n\varepsilon^2} \log \alpha(T_n, \theta, \varepsilon) \leq \frac{I(\theta)}{2},$$

$$\text{(3.3)} \quad \lim_{\varepsilon\to 0} \lim_{n\to\infty} -\frac{1}{n\varepsilon^2} \log \alpha(\hat{\theta}_n, \theta, \varepsilon) = \frac{I(\theta)}{2}.$$

(3.3) の等式を満たす推定量を，Bahadur の意味で，一次漸近有効な推定量と呼ぶ．したがって，MLE は，Bahadur の意味で，一次漸近有効となる．

(定理 3.1 の証明の概略)

この定理は、large deviation 理論と関連しており、次の Chernoff の定理 (Chernoff[3],Bahadur[1]) を用いて証明される.

$X_1, X_2, \cdots, X_n$ は独立同分布で、$E(X_i) \leq a$ の時、

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pr(\bar{X}_n \geq a) = -\log[\inf_{t\geq 0} E(e^{t(X-a)})].$$

方針としては、まず、(v) の解の一意性より、

$$l'(s \mid \theta') \geq 0 \quad if \quad \theta' \leq \hat{\theta}_n$$

$$l'(s \mid \theta') \leq 0 \quad if \quad \theta' \geq \hat{\theta}_n$$

したがって、

$$\Pr(\frac{1}{n}\Sigma l'(x_i \mid \theta + \varepsilon) > 0) = \Pr(\hat{\theta}_n - \theta > \varepsilon)$$

$$\Pr(\frac{1}{n}\Sigma l'(x_i \mid \theta - \varepsilon) < 0) = \Pr(\hat{\theta}_n - \theta < -\varepsilon)$$

ここで、各 $l'(x_i \mid \theta + \varepsilon), l'(x_i \mid \theta - \varepsilon)$ に Chernoff の定理を適応すると、

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pr(\hat{\theta}_n - \theta > \varepsilon) = -\log[\inf_{t\geq 0} \phi(t, \theta, \varepsilon)]$$

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pr(\hat{\theta}_n - \theta < -\varepsilon) = -\log[\inf_{t\leq 0} \phi(t, \theta, -\varepsilon)]$$

次に、(ii) の条件より剰余項の収束性がいえて、

$$E(l'(x \mid \theta + \varepsilon)) = -\varepsilon I + o(\varepsilon)$$

また、(iii) より、$\frac{\partial}{\partial t}\phi(t, \theta, \varepsilon) = 0$ の解 $\tau_\theta(\varepsilon)$ の一意性がいえる.
そして、(iii)(iv) より順序交換と展開を用いて、$\tau_\theta(\varepsilon) = \varepsilon + o(\varepsilon)$ となる.
最後に、$\inf_{t\geq 0} \phi(t, \theta, \varepsilon) = \phi(\tau_\theta(\varepsilon), \theta, \varepsilon)$ であることより、$-\log[\phi(\tau_\theta(\varepsilon), \theta, \varepsilon)]$ を ε^2 の項まで展開すればよい. $-\log[\inf_{t\leq 0} \phi(t, \theta, -\varepsilon)$ についても同様にすればよい.

分布が位置母数に対して対称である場合には，次のような 2 つの計算上有用な命題が，成立する.

命題 3.2 (Sievers[14])
標本平均 $\bar{X}_n$, メデイアン $X_{(\frac{1}{2})}$, MLE $\hat{\theta}_n$ の場合には，exponential rate は，次のように計算される.

$$\beta(\hat{\theta}_n, \theta, \varepsilon) = -\log[\inf_{t\geq 0} \phi(t, \theta, \varepsilon)],$$

$$\beta(\bar{X}_n, \theta, \varepsilon) = -\log[\inf_{t \geq 0} E\{e^{t(X-\theta-\varepsilon)}\}],$$

$$\beta(X_{(\frac{1}{2})}, \theta, \varepsilon) = -\frac{1}{2}\log\{4P(1-P)\}.$$

ここで, $P = \Pr(X > \theta + \varepsilon)$ である.

命題 3.3 (Fu[8])

$$\lim_{\varepsilon \to 0} -\frac{1}{\varepsilon^2}\beta(\bar{X}_n, \varepsilon, \theta) = \frac{1}{2\sigma^2},$$

$$\lim_{\varepsilon \to 0} -\frac{1}{\varepsilon^2}\beta(X_{(\frac{1}{2})}, \varepsilon, \theta) = 2f(0)^2.$$

ここで, σ^2 は, 分布の分散を表す.
したがって, 相対効率は, 漸近分散のものと一致する.

また, より一般的に, 次の命題は, exponential rate と漸近分散の密接な関係を示している.

命題 3.4 (Jurescŏvā and Kallenberg[11])
一致推定量 T_n が, ある正則条件の下で, $\sqrt{n}(T_n - \theta) \to_L N(0, v(\theta))$ であるなら,

$$\lim_{\varepsilon \to 0}\lim_{n \to \infty} -\frac{1}{n\varepsilon^2}\log\Pr(| T_n - \theta | > \varepsilon) = \frac{1}{2v(\theta)}$$

が成立する.

Bahadur の意味で一次漸近有効な推定量の中で, 区別をつけるためには, より高次の有効性を考えなければならない. その時, 次の定理が成立する.

定理 3.5 (Fu[9],Fu[10])
(1) 正則条件 (RC) において, (ii) と (iv) を, 次の様にかえると, (3.4) が成立する.
(ii)' すべての $\theta \in \Theta$ に対して, 適当な $N_{\theta,\delta}$ と, $A_4(x, \theta)$ が存在して, すべての $\theta', \theta'' \in N_{\theta,\delta}$ に対して, $E(A_4^2) < \infty, | l^{(4)}(x \mid \theta'') - l^{(4)}(x \mid \theta') | \leq A_4 \mid \theta'' - \theta' |$
(iv)' 偏導関数 $(\frac{\partial}{\partial t})^4(\frac{\partial}{\partial \theta})^4\phi(t, \theta, \varepsilon)$ が, $(t, \theta) \in (-u, u) \times (-w, w)$ に対して, 連続である.

$$(3.4) \qquad \lim_{\varepsilon \to 0}\frac{1}{\varepsilon^i}\{B(\theta, \varepsilon) - \beta(\hat{\theta}_n, \theta, \varepsilon)\} = \begin{cases} 0 & \text{if} \quad i = 1, 2, 3 \\ \frac{1}{8}I^2\gamma^2 & \text{if} \quad i = 4 \end{cases}$$

(2) 分布が位置母数に対して対称な場合，位置不変一致推定量 T_n の中で

$$(3.5) \qquad \lim_{\varepsilon\to 0}\frac{1}{\varepsilon^4}\{B(\theta,\varepsilon)-\beta(T_n,\theta,\varepsilon)\}\geq\frac{1}{8}I^2\gamma^2.$$

$T_n=\hat{\theta}_n$ の時，等号が成立する．

(3.5) で等号が成立する推定量を，Bahadur の意味で二次漸近有効であるという．したがって，MLE は，二次漸近有効である．

注意 3.1 exponential rate による比較での限界とバイアスの調整

$\lim_{n\to\infty}C_n=1$ となる数列 C_n において，一般に，T_n と C_nT_n は同じ exponential rate をもち，Bahadur の意味では，推定量の良さを区別することはできない．

例 3.1　一様分布 $U(0,\theta)$ において，MLE は、$X_{(n)}$ で、不偏推定量は、$\frac{n+1}{n}X_{(n)}$ である．この 2 つの推定量において、n が十分に大きいとき、

$$\alpha(X_n,\theta,\varepsilon)=(\frac{\theta-\varepsilon}{\theta})^n$$

$$\alpha(\frac{n+1}{n}X_n,\theta,\varepsilon)=(\frac{n}{n+1})^n(\frac{\theta-\varepsilon}{\theta})^n$$

となり、exponential rate は共に、$\log\frac{\theta}{\theta-\varepsilon}$ となる．これは Bahadur bound と一致している．ただし，集中度　$\alpha(\cdot,\theta,\varepsilon)$ の値では，$\frac{n+1}{n}X_{(n)}$ の方が良い．

一般に，バイアスを調整した MLE を，最適とみなす．この考え方は，定理 2.1 での Rao-Efron による平均二乗損失の下での主張と一致する．

注意 3.2 exponential rate と，漸近分散との対比

命題 3.3 と命題 3.4 より，この 2 つの量での有効性は，ある場合には全く同じになる．また，一般には，前者の方が，得るのが難しく，有用性に欠けるように思える．しかし，正則条件が，満たされない時には，次の例の様に，前者の場合のみ，推定量の良さを比べる事が出来る．

例 3.2　(Juresčovā and kallenberg[11])
$X_1,X_2,\cdots,X_n$ は，$N(\theta,1)$ の独立同分布に従っている．

$$Y_n=\frac{1}{n-1}\sum_{i=2}^{n}X_i,\qquad Z_n=\sum_{i=2}^{n}(X_i-Y_n)^2,\qquad n=2,3,\cdots,$$

$$H_n = \begin{cases} 0 & \text{if} \quad Z_n \le \chi^2_{n-2,1-\frac{1}{n}} \\ 1 & \text{otherwise} \end{cases} \qquad n = 2, 3, \cdots$$

と定義する．ここで，$\chi^2_{k,\alpha}$ は，自由度 k のχ^2 分布の α 分位点である．そして，推定量 T_n を，

$$T_n = (1 - H_n)Y_n + H_n X_1 \qquad n = 2, 3, \cdots$$

と定義すると，

$$\sqrt{n}(T_n - \theta) \to_L N(0,1) \qquad n \to \infty$$

となり，$\bar{X}_n$ と同じ漸近分布を持つことがわかる．しかし，この T_n において

$$\alpha(T_n, \theta, \varepsilon) = \Pr(H_n = 1, | X_1 - \theta | > \varepsilon) + \Pr(H_n = 0, | Y_n - \theta | > \varepsilon)$$

となり、Y_n , Z_n , H_n の独立性を考えると、

$$\alpha(T_n, \theta, \varepsilon) = \frac{2}{n}\Phi(-\varepsilon) + 2(1 - \frac{1}{n})\Phi(-\varepsilon\sqrt{n-1}).$$

ここで，$\Phi(\cdot)$ は，標準正規分布の分布関数を表す．故に、

$$\alpha(T_n, \theta, \varepsilon) \ge \frac{2}{n}\Phi(-\varepsilon)$$

$$\beta(T_n, \theta, \varepsilon) \le \lim_{n\to\infty} -\frac{1}{n}\log\frac{2}{n}\Phi(-\varepsilon) = 0$$

従って，exponential rate は，0 となる．このことより，$\bar{X}_n$ と T_n とは，漸近分散では区別できないが，exponential rate では，$\bar{X}_n$ の方が，優れている事がわかる．

注意 3.3 Cauchy 分布の有効推定量
Cauchy 分布の位置母数の推定においては，尤度方程式が多解をもつことより (RC)(v) を満たさず定理 3.1 を適応できないが，この場合にも，(3.3) が成立し、MLE が Bahadur の意味で一次有効であることが示されている．(Bai and Fu[2])

注意 3.4 超有効推定量の影響
Bahadur の有効性は，漸近分散の場合と違って，超有効推定量の害を除くことができる．

例 3.3 $N(\theta, 1)$ の推定における Hodges の超有効推定量

$$\tilde{\theta}_n = \begin{cases} \bar{X}_n & \text{if} \quad | \bar{X}_n | \ge n^{-\frac{1}{4}} \\ \alpha\bar{X}_n & \text{otherwise} \end{cases}$$

とすると $(0<\alpha<1)$,

$$\sqrt{n}(\hat{\theta}_n-\theta)\to_L N(0,1),$$

$$\sqrt{n}(\tilde{\theta}_n-\theta)\to_L \begin{cases} N(0,1) & \text{if} \quad \theta\neq 0 \\ N(0,\alpha^2) & \text{if} \quad \theta=0 \end{cases}$$

となり,$\theta=0$ で超有効点をもつ.しかし,Bahadur の有効性に関しては,すべての θ に対して

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}-\frac{1}{n\varepsilon^2}\log \Pr(|\tilde{\theta}_n-\theta|>\varepsilon)=\frac{1}{2}$$

となり,超有効点をもたない.

参考文献

[1] Bahadur, R. R. (1971)*Some limit theorems in statistics.,* Regional Conference Series in Applied Mathematics, 4. SIAM, Philadelphia.

[2] Bai, Z. D. and Fu, J. C. (1987) On the maximum-likelihood estimator for the location parameter of a Cauchy distribution. *Canadian J. of Stat.,* **15**, 137–146.

[3] Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.,* **23**, 493–507.

[4] Efron, B. (1975) Defining the curvature of a statisticar problem. *Ann. Stat.,* **3**, 1189–1242.

[5] Fisher, R. A. (1922) On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London, Ser. A.,* **222**, 309–368.

[6] Fisher, R. A. (1925) Theory of statistical estimation. *Proc. Cambridge Philos. Soc.,* **22**, 700–725.

[7] Fu, J. C. (1973) On a theorem of Bahadur on the rate of convergence of point estimators. *Ann. stat.,* **1**, 745–749.

[8] Fu, J. C. (1975) The rate of convergence of consistent point estimators. *Ann. Stat.,* **3**, 234–240.

[9] Fu, J. C. (1982) Large sample point estimation: A large deviation theory approach. *Ann. Stat.* , **10**, 762–771.

[10] Fu, J. C. (1985) On exponential rates of likelihood ratio estimators for location parameters. *Stat. & Prob. Let.*, **3**, 101–105.

[11] Jurečkovā, J. and Kallenberg, W. C. M. (1987) On local inaccuracy rates and asymptotic variances. *Stat. & Dec.*, **5**, 139–158.

[12] Rao, C, R. (1961) Asymptotic efficiency and limiting information. *Proc. fifth Berk. Sympo.*, **1**, 531–545.

[13] Rao, C. R. (1963) Criteria of estimation in large sample. *Sakhyā*, **25**, 189–206.

[14] Sievers, G. L. (1978) Estimates of location: A large deviation comparison. *Ann. Stat.*, **6**, 610–618.